PU Learning for Matrix Completion

Cho-Jui Hsieh, *et al*. ICML, 2015.

Matrix Completion

- Example: movie recommendation
- Given a set Ω and the values M_{Ω}, how to predict other elements?

movies

users

		6.8			1.9	
		7.9				4.4
5.1			?		2.3	
6.8						5.3
				9.3		
	8.8			8.0		
5.1			7.7		1.9	

Matrix Completion

- Assumption: the underlying matrix M is low rank.
- Recover M by solving

$$\min_{\|X\|_* \leq t} \sum_{i,j \in \Omega} (X_{ij} - M_{ij})^2$$

0.9	2.7
2.1	1.8
2.9	1.1
3.4	1.7
1.6	1.6
2.1	0.7
1.9	1.6

*

1	3.3	2.2	1.8	2.8	0.6	0.8
2	2.7	1.8	2.7	3.0	0.5	1.5

=

		6.8			1.9	
		7.9				4.4
5.1			8.2		2.3	
6.8						5.3
				9.3		
	8.8			8.0		
5.1			7.7		1.9	

PU Matrix Completion

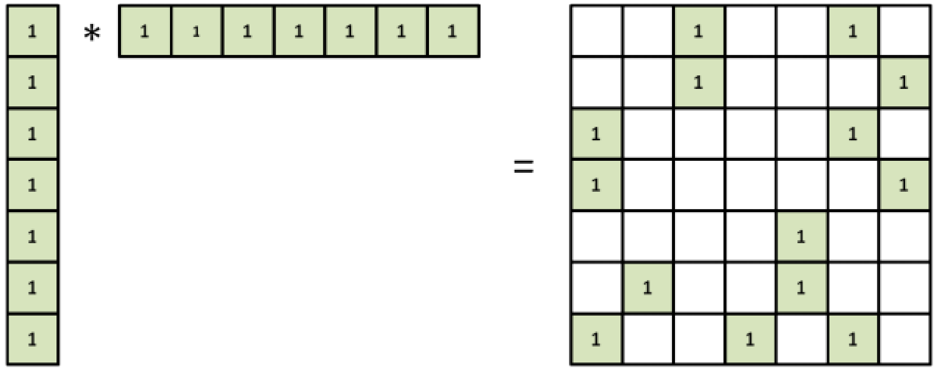
- All the observed entries are 1's.
- Example :
  - Link prediction in social networks (only friend relationships)
  - Product recommendation

users (columns) × users (rows)

		1			1	
		1				1
1	?	?	?	?	1	?
1						1
				1		
	1			1		
1			1		1	

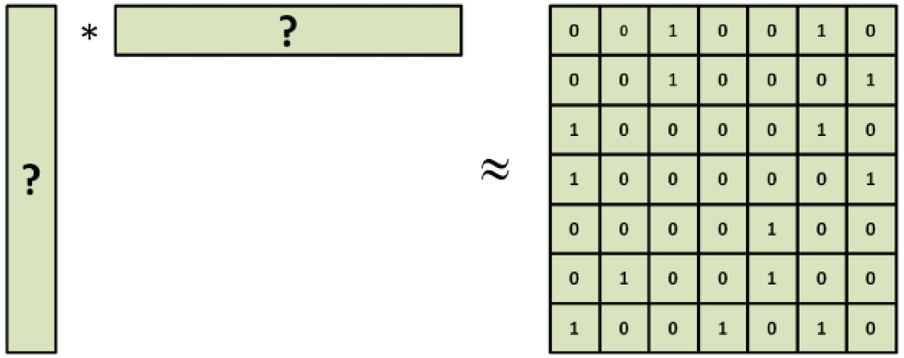
Problem

- Minimizing the loss on the observed 1's.



trivial solution

- Treat all the missing entries as zeroes, and minimizing the loss on all the entries.



99% elements are zero → tend to fit zeroes instead of ones

Outline

- Shifted Matrix Completion for Non-deterministic Setting
- Biased Matrix Completion for Deterministic Setting
- PU Inductive Matrix Completion
- Experiment

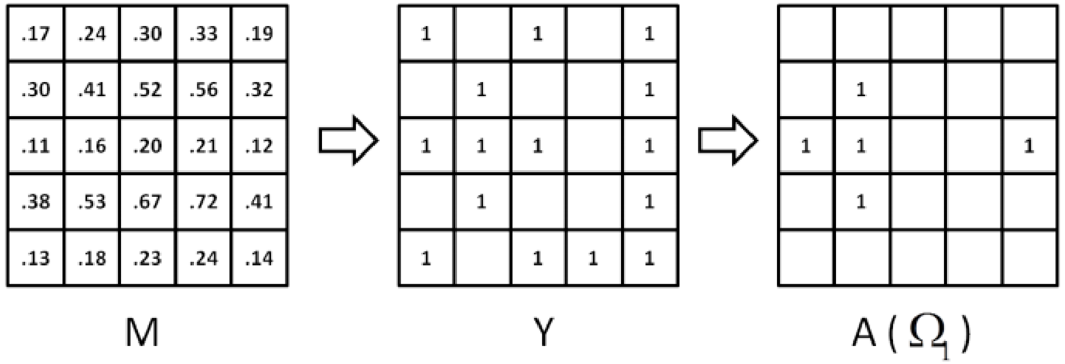
Non-deterministic Setting

- $M_{ij} \in [0,1]$, M is low rank.
- The generating process: M(underlying) $\rightarrow$ Y(0-1 matrix) $\rightarrow$ Ω_1
- An underlying 0-1 matrix Y is generated by

$$Y_{ij} = \begin{cases} 1 & \text{with } M_{ij} \\ 0 & \text{with } 1 - M_{ij} \end{cases}$$

- Ω_1 sampled from $\{(i,j) | Y_{ij} = 1\}$, the sample rate is $1 - \rho = |\Omega_1| / \|Y\|_0$

recover the underlying matrix M



Unbiased Estimator

$\rho = 1 - \bar{s}/s$ is the noise rate of flipping a 1 to 0

$1-\rho$ is the sampling rate to obtain Ω_1 from Y

$A \in \{0,1\}^{m \times n}$: observations

$$\begin{cases} A_{\Omega_1} = 1 \\ A_{ij} = 0 \text{ for all } (i,j) \notin \Omega_1 \end{cases}$$

$$P(A_{ij}=1) = M_{ij}(1-\rho), \; P(A_{ij}=0) = 1 - M_{ij}(1-\rho)$$

M

.17	.24	.30	.33	.19
.30	.41	.52	.56	.32
.11	.16	.20	.21	.12
.38	.53	.67	.72	.41
.13	.18	.23	.24	.14

⇨

Y

1		1		1
	1			1
1	1	1		1
	1			1
1		1	1	1

$s = \#1$

⇨

A (Ω_1)

	1			
1	1			1
	1			

$\bar{s} = |\Omega_1|$

$$\ell(X_{ij}, M_{ij}) = (X_{ij} - M_{ij})^2$$

⇨

$$\tilde{\ell}(X_{ij}, A_{ij}) = \begin{cases} \frac{(X_{ij}-1)^2 - \rho X_{ij}^2}{1-\rho} & \text{if } A_{ij} = 1, \\ X_{ij}^2 & \text{if } A_{ij} = 0. \end{cases}$$

Shifted Matrix Completion

$$\min_X \sum_{i,j} \tilde{\ell}(X_{ij}, A_{ij}) \ \text{ such that } \|X\|_* \leq t, 0 \leq X_{ij} \leq 1 \ \forall (i,j).$$

$$\text{where } \ \tilde{\ell}(X_{ij}, A_{ij}) = \begin{cases} \frac{(X_{ij}-1)^2 - \rho X_{ij}^2}{1-\rho} & \text{if } A_{ij} = 1, \\ X_{ij}^2 & \text{if } A_{ij} = 0. \end{cases}$$

$$\textit{For any } X \in \mathbb{R}^{m \times n}, \ \frac{1}{mn} E\big[\textstyle\sum_{i,j} (X_{ij} - Y_{ij})^2\big] = \frac{1}{mn} E\big[\textstyle\sum_{i,j} \tilde{\ell}(X_{ij}, A_{ij})\big]$$

$$\min_X \sum_{i,j} \tilde{\ell}(X_{ij}, A_{ij}) + \lambda \|X\|_*, \ \text{ such that } 0 \leq X_{ij} \leq 1 \ \forall (i,j)$$

$$\hat{X} = \underset{X}{\operatorname{argmin}} \sum_{i,j:A_{ij}=1} \left(X_{ij} - \frac{1}{1-\rho} \right)^2 + \sum_{i,j:A_{ij}=0} X_{ij}^2 + \lambda \|X\|_*$$
$$\text{s.t. } 0 \leq X_{ij} \leq 1 \ \forall (i,j). \tag{7}$$

Deterministic Setting

- $M_{ij} \in [0,1]$
- With threshold $q \in [0,1]$: $Y_{ij} = \begin{cases} 1 & \text{with } M_{ij} > q \\ 0 & \text{with } M_{ij} \le q \end{cases}$

- Ω_1 sampled from $\{(i,j) | Y_{ij} = 1\}$
- The goal is to recover Y

.17	.24	.30	.33	.19
.30	.41	.52	.56	.32
.11	.16	.20	.21	.12
.38	.53	.67	.72	.41
.13	.18	.23	.24	.14

M

⇨

		1	1	
	1	1	1	1
1	1	1	1	1

Y

⇨

	1		1	
	1		1	

$A(\Omega_1)$

Label-dependent Loss

$$\text{thr}(x) = 1 \text{ if } x > q$$
$$\text{thr}(x) = 0 \text{ if } x \leq q$$

Recovery error: $$R(X) = \frac{1}{mn} \sum_{i,j} 1_{\text{thr}(X_{ij}) \neq Y_{ij}}$$

Label-dependent error: $$U_\alpha(x, a) = (1-\alpha) 1_{\text{thr}(x)=1} 1_{a=0} + \alpha 1_{\text{thr}(x)=-1} 1_{a=1}$$

α-weighted expected error: $$R_{\alpha,\rho}(X) = E\Big[\sum_{i,j} U_\alpha(X_{ij}, A_{ij})\Big]$$

Lemma 2. *For the choice* $\alpha^* = \frac{1+\rho}{2}$ *and* $\beta = \frac{1+\rho}{2}$, *there exists a constant* b *that is independent of* X *such that, for any matrix* X, $R_{\alpha^*,\rho}(X) = \beta R(X) + b$.

minimizing the α-weighed expected error in the partially observed situation = minimizing the true recovery error R

Biased Matrix Completion

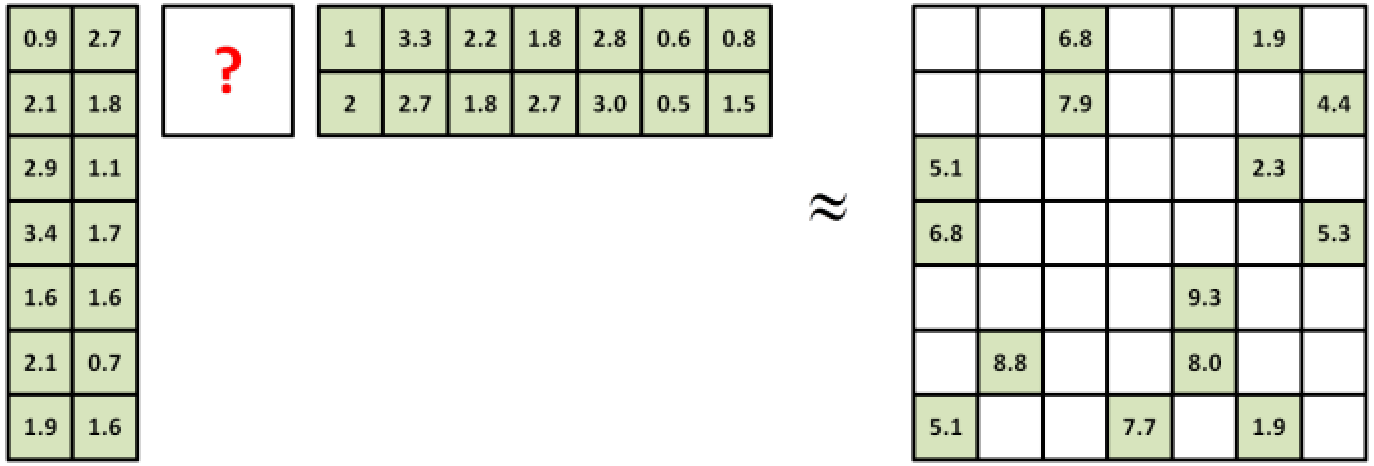
$$\ell_\alpha(x,a) = \alpha 1_{a=1}\ell(x,1) + (1-\alpha)1_{a=0}\ell(x,0)$$

$$\hat{X} = \underset{X:\|X\|_*\leq t}{\text{argmin}} \sum_{i,j} \ell_\alpha(X_{ij}, A_{ij}) = \underset{X:\|X\|_*\leq t}{\text{argmin}}\ \alpha \sum_{i,j:A_{ij}=1} (X_{ij}-1)^2 + (1-\alpha) \sum_{i,j:A_{ij}=0} X_{ij}^2$$

PU Inductive Matrix Completion

- Input: partially observed matrix A_Ω and features $F_u, F_v \in R^{n\times d}$ associated with rows/columns.
- Recover the underlying matrix by solving

$$\min_{D\in\mathbb{R}^{d\times d}} \sum_{i,j\in\Omega} (A_{ij} - (F_u D F_v^T)_{ij})^2 + \lambda \|D\|_*.$$



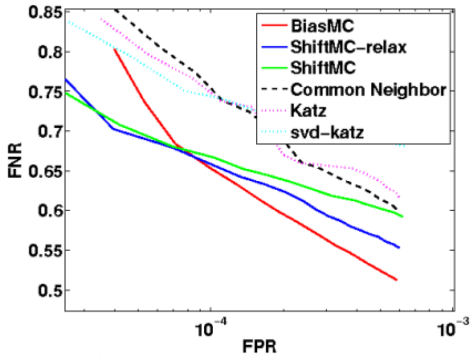
Experiment - Link Prediction

ca-GrQc (4,158 nodes and 26,850 edges)

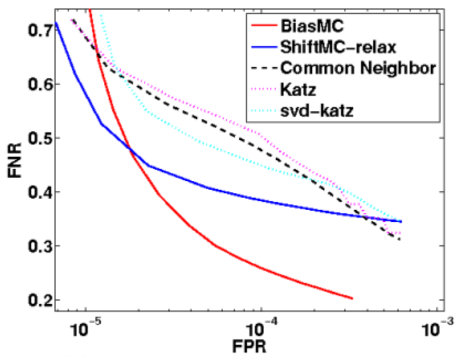
ca-HepPh (11,204 nodes and 235,368 edges)

LiveJournal (1,770,961 nodes, $|\Omega_{train}|$= 83,663,478 and $|\Omega_{test}|$= 2,055,288)

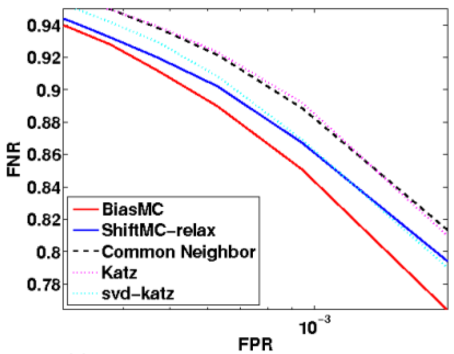
MySpace (2,137,264 nodes, $|\Omega_{train}|$= 90,333,122 and $|\Omega_{test}|$= 1,315,594)



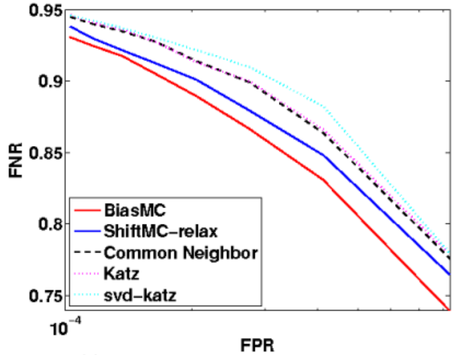
(c) FPR-FNR on ca-GrQc dataset.



(d) FPR-FNR on ca-HepPh dataset.



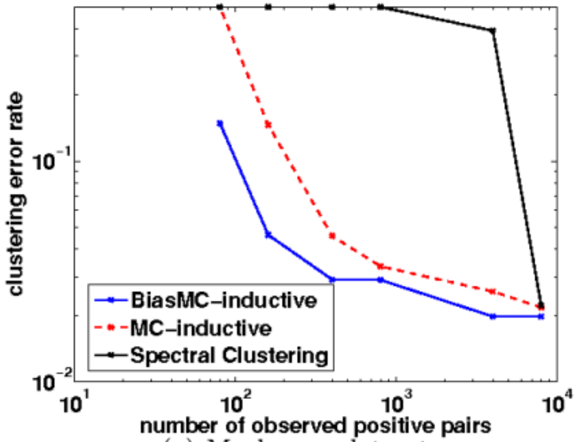
(e) FPR-FNR on LiveJournal dataset.



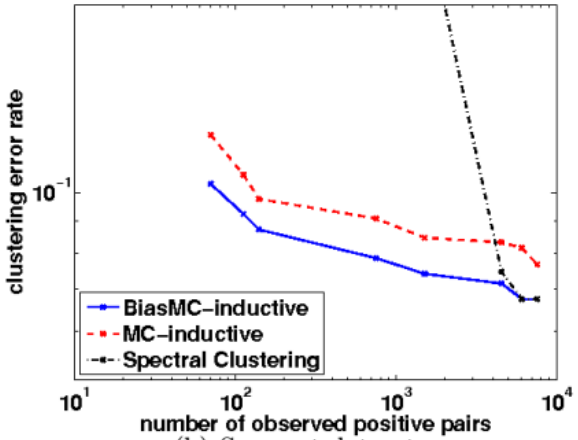
(f) FPR-FNR on MySpace dataset.

Experiment - Inductive matrix completion

In semi-supervised clustering problems, we are given n samples with features $\{\boldsymbol{x}_i\}_{i=1}^n$ and pairwise relationships $A \in \mathbb{R}^{n \times n}$, where $A_{ij} = 1$ if two samples are in the same cluster, $A_{ij} = -1$ if they are in different clusters, and $A_{ij} = 0$ if the relationship is unobserved.



(a) Mushroom dataset.



(b) Segment dataset.